Bibliography:

1. Apostolico, A. "Optimal Parallel Detection of Squares in Strings", Algorithmica 8 (4): 285-319 (1992)

2. Abascal, F. and Valencia, A. (2002). "Clustering of Proximal Sequence Space for the Identification of Protein Families," *Bioinformatics*, 2002

3. Benson, G., Waterman, M. S. "A method for fast database search for all knucleotide repeats", *Nucleic Acids Research*, 22:4828- 4836, 1994

4. Benson, G. "Tandem repeat finder: a program to analyze DNA sequences", *Nucleic Acids Research*, 27: 573-580, 1999

5. Benson, G. "Sequence Alignment with Tandem Duplication", *Journal of Computational Biology 4*(3): 351-368 (1997)

6. Chang, W. I. and Lawler, W.L. "Approximate string matching in sublinear expected time Proc*", Proc. 31st Annual IEEE Symposium on Foundations of Computer Science*, 116-124, 1990

7. Chiang, J.H., Yu, H.C., and Hsu, H.J. (2004). "GIS: A Biomedical Text-mining System for Gene Information Discovery," Bioinformatics, 20(1):120-1.

8. Edwards, A., Hammond, H., Jin, L., Caskey. C. and Chakraborty, R. “Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups”, *Genomics*, 12: 241-253, 1992

9. Fischetti, V., Landau, G., Schmidt, J. and Sellers, P, “Identifying periodic occurrences of a template with applications to protein structure”, *Third Annual Symposium on Combinatorial Pattern Matching conference*, 111-120, 1992

10. Fu, Y,H., Pizzuti,A., Fenwick, J.,King, R.G.Jr. and Rajnarayan, S., Dunne, P. W., Dubel, J., Nasser, G.A., Ashizawa, T., DeJong, P., Wieringa, B. Korneluk, R., Perryman, M.B., Epstein, H. F. and Caskey, “an unstable triplet repeat in a gene related to myotonic muscular dystrophy”, Science, 255: 1256-1258, 1992

11. Edgar, R., Domrachev, M., and Lash, A.E. (2002). “Gene Expression Omnibus: NCBI Gene Expression and Hybridization Array Data Repository,” Nucleic Acids Res., 30(1):207-210.

12. Hamada, H., Seidman, M., Howard, B. and Gorman, “Enhanced gene expression by the poly(dT-dG) poly(dC-dA) sequence”, *Molecular and Cellular Biology*, 4: 2622-2630, 1984

13. Han, J., Jamil, H. M., Lu, Y., Chen, L., Liao, Y. and Pei, J., “DNA-Miner: A System Prototype for Mining DNA Sequences”, *In the Proceedings of the ACM SIGMOD*

International Conference on the Management of Data, May 21-24, 2001, Santa Barbara, CA, USA.

14. Han, J., Pei, J., Mortazavi-Asl, B., Chen, Q., Dayal, U. and Hsu, M.-C. "Frequent pattern projected sequential pattern mining", *In proc. 2000 Int. Conf. Knowledge Discovery and Data mining (KDD)* 355-359, Boston, MA, August 2000

15. Hellman, LL. Steen, M., Sundvall, M. and Pettersson, U. "A rapidly evolving region in the immunoglobulin heavy chain loci of rat and mouse: postulated role of (dCdA) n (dG-dT)n sequences", *Gene*, 68: 93-100, 1888

16. Hirschberg, D.S. Comm, "A linear space algorithm for computing maximal common subsequences," *Communications of the ACM*, 18, 341-343, 1975

17. Huntington's Disease Collaborative Research Group. "A novel gene containing a trinucleotide repeat that is expanded an unstable on Huntington's disease chromosomes", *Cell*, 72: 971-983, 1993

18. Karlin, S., Morris, M., Ghandour, G. and Leung, M.-Y. "Efficient algorithms for molecular sequence analysis", *Proc. Natl. Acad. Sci.* USA, 85: 841-845, 1988

19. Leung, M. Y., Blaisdell, B.E., Burge, C. Karlin, S. “An efficient algorithm for identifying matches with errors in multiple long molecular sequences”, *J. Mol. Biol*., 221: 1367-1378, 1991

20. Karp, R. M., Miller, R. E. and Rosenberg, A. L. “Rapid identification of repeated patterns in strings, trees and arrays”, *In Proc. 4th Annu. ACM Symp. Theory of Computing*, 125-136, 1972

21. V. I. Levenshtein, “Binary Codes Capable of Correcting Deletions, Insertions and Reversals,” *Soviet Phys. Dokl*, Vol. 10, pp. 707-710, 1966

22. Landau, G. and Schmidt, “An algorithm for Approximate Tandem repeats”, *In Proc. 4th Annual Symp. on Combinatorial Pattern Matching, Lecture Notes in Computer Science*, Vol 648: 120-133, 1993

23. La Spada, A., Wilson, E., Lubahn, D., Harding, A. and Fischbeck, K, “Androgen receptor gene mutations in X-linked spinal and bulbar muscular atroply,” *Nature*, 352: 77-79, 1991

24. Lu, Q., Wallrath, L., Granok, H. and Elgin, S., “(CT)n (GA)n repeats and heat shock elements have distinct roles in chromatin structure and transcriptional activation of the Drosophila hsp26 gene”, *Mol. Cell. Biol*. 13: 2802-2814, 1993

25. Milosavljevic, A. and Jurka, “Discovery simple DNA sequences by the algorithmic significance method”, *Computer applications in the bioinformatics*, 9: 407- 411, 199

26. M.G. Main and R.J. Lorentz, “An O(nlogn) algorithm for finding all repetitions in a string,” *J. of Algorithms*, Vol 5, 422-432, 1984

27. M.G. Main and R.J. Lorentz “Linear time recognition of square free strings,” A. Apostolico and Z. Galil (deitors), *Combinatorial Algorithms on Words*, NATO ASI Series, F: Computer and System Science, Vol. 12, Springer-Verlag, 272-278

28. Myers, E.W. and Miller, “Sequence comparison with concave weighting functions”. *Bull Math Biol*. 50: 97-120 ,1988

29. Mannila, H., Toivonen, H., and Verkamo, A. I. “Discovery of frequent episodes in event sequences”, in Data Mining and Knowledge Discovery, 1: 259-289, 1997

30. Palakal, M., Stephens, M., Mukhopadhyay, S., Raje, R., and Rhodes, S.J. (2002). “A Multilevel Text Mining Method to Extract Biological Relationships,” IEEE CSB:97-108.

31. Pei, J., Han, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U., and Hsu, M.-C. “Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth”, *In proc. 2001 Int. Conf. Data Engineering (ICDE)*, Heidelberg, Germany, April 2001

32. Pardue, M., Lowenhaupt, K., Rich, A. and Nordheim, A., "(dC-dA)n (dGdT) n sequence have evolutionarily conserved chromosomal locations in Drosophila with implications for roles in chromosome structure and function", the *EMBO Journal*, 6: 1781-1789, 1987

33. Richards, R., Holman, K., Yu,S. and Southerland, G., "Fragile X syndrome unstable element, p(CCG)n, and other simple tandem repeat sequences are binding sites for specific nuclear proteins", *Hum. Mol. Genet*., 2: 1429-1435, 1993

34. Rival, E., Delgrange, O., Delahaye, J.P., Dauchet, M., Delorme, M. Henaut, A. and Ollivier, E., "Detection of significant patterns by compression algorithms: the case of approximate tandem repeats in DNA sequences. *CABIOS*, 13: 131-136, 1997

35. M.-F. Sagot et E. W. Myers, "Identifying satellites in nucleic acid sequences, RECOMB'98", (S. Istrail, P. Pevzner et M. Waterman, éds.) New York, USA, *Proceedings of the Second Annual International Conference on Computational Molecular Biology*, pages 234-242, ACM Press, 1998

36. Salwinski, L, Miller, C.S., Smith, A.J., Pettit, F.K., Bowie, J.U., and Eisenberg, D. (2004).
"The Database of Interacting Proteins: 2004 Update," Nucl. Acids. Res., 32:D449-D451.

37. Smith, T.F. and Waterman.M.S. "Identification of common molecular subsequences", *J. Mol.Biol*., 147: 195-197, 1981

38. Srikant R. and Agrawal, R. “Mining sequential patterns: generations and performance improvements”, *In proc. 5th Int. Conf. Extending Database Technology (EDTB)* 3-17, Avignon, France, March, 1996

39. Schnase, J.H., Lane, M.A., Bowker, G.C., Star, S.L. and Silberschatz, A. “Building the next-generation biological-information infrastructure”, *Nature and Human Society: The Quest for a Sustainable World,* Washington: National Academy Press, 2000

40. Tanay, A., Sharan, R., and Shamir, R. (2002). Discovering statistically signi_cant biclusters in gene expression data. Bioinformatics, 18(Suppl1):S136{S144.

41. TheGenomic data Mine, Lorraine Tanabe National Center for Biotechnology Information, Computational Biology Branch, National Library of Medicine, Bethesda, 2000

42. Verkerk,A., Pieretti, M., Sutcliffe, J. Fu, Y., Kul,D., Pizzuti, A., Reiner,O., Rechards,S., Victoria, M., Zhang, F., Eussen, B., van Ommen,G., Blonden, A., Riggins, G., Chastain, J., Kunst, C., Galjaard, H., Caskey, C., Nelson, D., Oostra, B. and Warren, S. “Identification of gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome”, *Cell,* 65, 05-914, 1991

43. Wang, J., Shapiro,B., Shasha,D., *“Pattern Discovery in Biomolecualr Data”,* New York Oxford, Oxford University Press, 1999